An evaluation of machine intelligence tools to diagnose genetic diseases in critically ill infants

Michelle Clark PhD
Statistical Scientist, Rady Children's Institute for Genomic Medicine

Rady Children's Institute
Genomic Medicine

Background

Rady Children's Institute Genomic Medicine

55% disease of unknown etiology

Search for etiological diagnosis

Interim empirical treatment

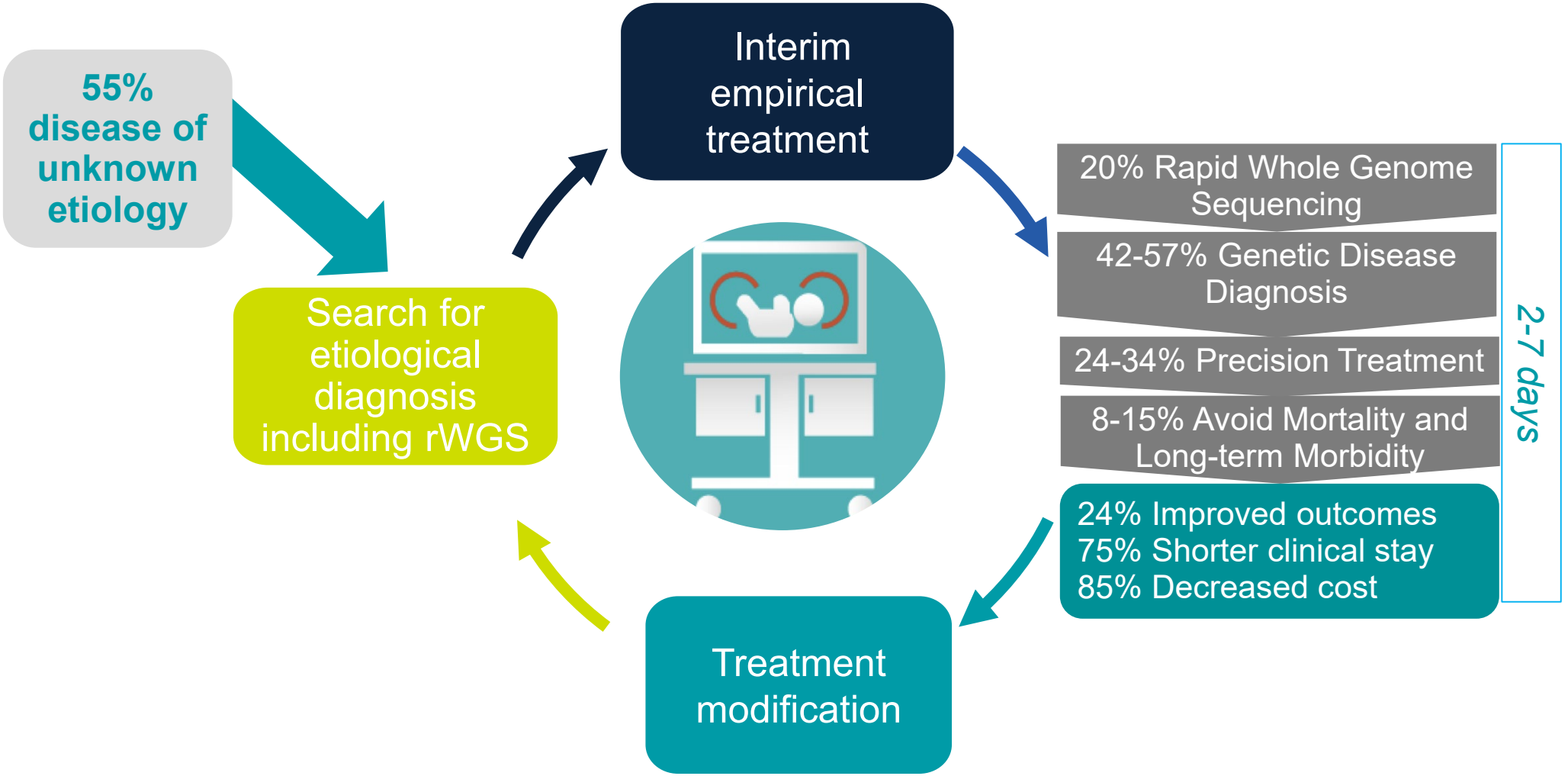
Improvement or worsening

- Discharged home undiagnosed
- Delayed diagnosis: Anxiety, suffering
- Unnecessary morbidity and mortality
- Delayed palliative care
- Unnecessary cost

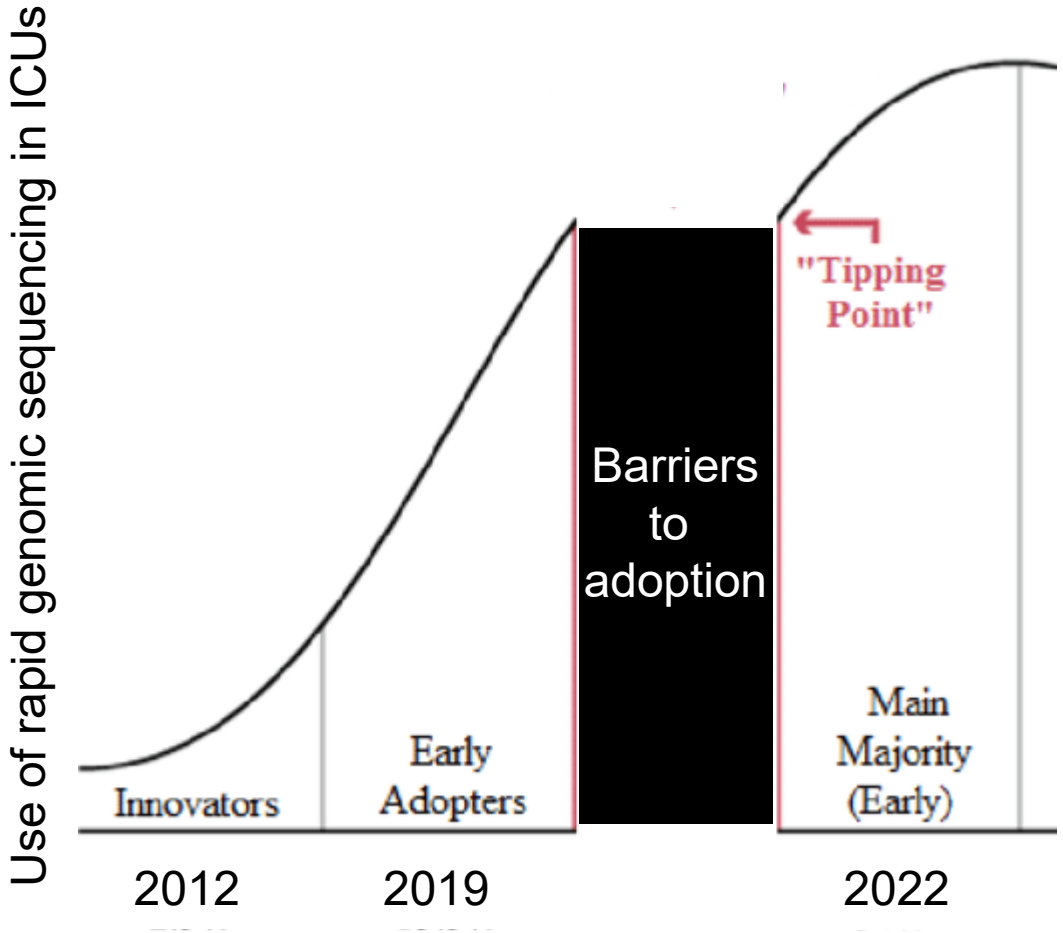
Treatment modification

Diagnostic and clinical utility of rapid whole genome sequencing

Rady Children's Institute Genomic Medicine



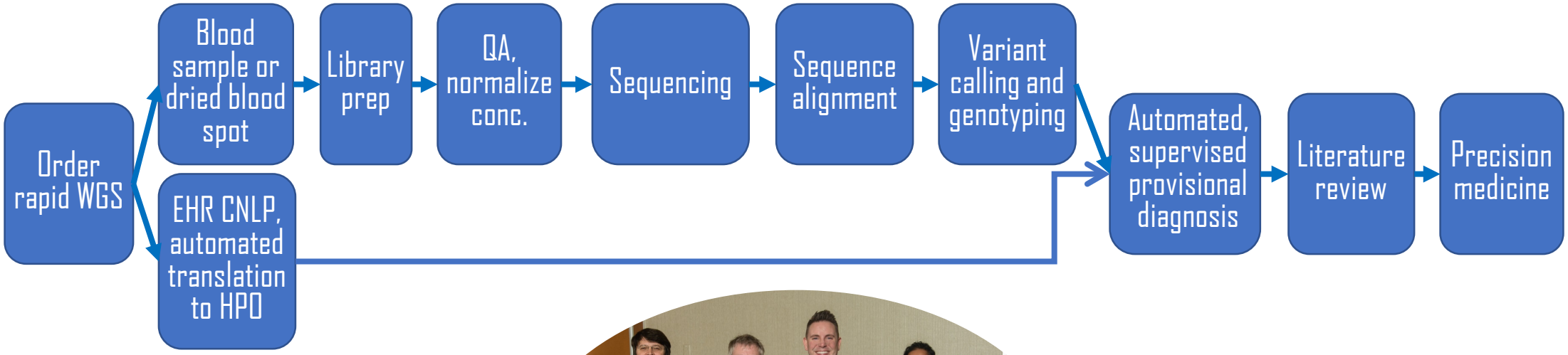
Barriers to broad adoption



- **Capital & labor intensity of rapid genomic sequencing**
  - **Shortage of expert medical geneticists, genetic counselors**
  - **Not scalable**
  - **Delays rapid changes in patient care**
- **Unfamiliarity with rapid genomic medicine**
  - **13,000 genetic diseases – most of them too rare to have been seen before by pediatricians**
- **Insufficient evidence of efficacy**
  - **Delayed authorization, failure of reimbursement**
- **Many genetic diseases lack effective treatments**
  - **Most treatments have not undergone rigorous testing**

Solution: automated diagnostic platform using machine intelligence

Rady Children's Institute Genomic Medicine



Time from blood draw to provisional diagnosis: 19.5 hours

Automated deep phenotyping

Rady Children's Institute Genomic Medicine

Automated variant interpretation

Rady Children's Institute Genomic Medicine

Evaluation of the automated diagnostic platform

Rady Children's Institute Genomic Medicine

1. **Retrospective study – 84 children**
2. **Timed study – 10 children**
3. **Reanalysis study – 48 children**
4. **Prospective study – 50 children**

1. Performance in a retrospective cohort: 99% precision, 97% recall

Rady Children's Institute Genomic Medicine

- **95 children with 97 genetic diseases diagnosed manually by rapid whole genome or whole exome sequencing with manually extracted phenotypes and manual interpretation**
- **Excluded incidental findings**
- **99% precision (93 of 94)**
- **97% recall (94 of 97)**

2. Timed study: 100% precision/recall
Mean time savings: 22hrs

Rady Children's Institute Genomic Medicine

Use Type	Retrospective Patients				Prospective Patients													
Subject ID	263		6124	3003	6194		290		352		362		374		7052		412	
Age	8 days		14 years	1 year	5 days		3 days		7 weeks		4 weeks		2 days		17 months		3 days	
Sex	♀		♂	♀	♀		♂		♀		♂		♂		♂		♂	
Abbreviated Presentation	Neonatal seizures		Rhabdo-myolysis	Dystonia, Dev. delay	Hypoglycemia, seizures		Pulmonary hemorrhage, PPHN		Diabetic ketoacidosis		Neonatal seizures		HIE, anemia		Pseudomonal septic shock		Neonatal seizures	
Method	Auto.	Auto.	Auto.	Auto.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.
Number of Phenotypic Features	51		115	148	14	2	257	4	103	4	65	1	112	6	124	3	33	1
Molecular Diagnosis	Early Infantile Epileptic Encephalopathy 7		Glycogen Storage Disease V	Dopa-Responsive Dystonia	None		None		Permanent neonatal diabetes mellitus		None		None		X-linked agamma-globulinemia 1		Benign familial neonatal seizures 1	
Gene and Causative Variant(s)	*KCNQ2* c.727C>G		*PYGM* c.2262delA c.1726C>T	*TH* c.785C>G c.541C>T	None		None		*INS c.26C>G*		None		None		*BTK* c.974+2T>C		*KCNQ2* c.1051C>G	
Sample/Library Prep (hours)	3:20	2:55	2:24	2:22	2:10	23:54	2:12	22:05	2:13	15:42	2:31	18:30	3:30	10:10	4:30	12:10	3:05	23:50
NovaSeq Loading (hours)	0:20	0:17	0:16	0:20	1:38	0:20	0:29	0:22	0:30	0:53	0:15	2:30	0:45	0:35	1:00	1:00	0:20	0:53
2x101 nt Sequencing (hours)	15:36	15:31	15:34	15:27	15:26	24:13	15:25	24:08	15:21	22:44	15:17	33:36	15:17	21:07	15:19	22:46	15:58	21:00
1^0 & 2^0 Analysis (hours)	1:03	1:02	0:59	0:59	1:07	3:05	1:00	1:57	1:01	2:30	1:02	2:30	1:02	2:30	1:09	2:25	1:24	2:24
3^0 Analysis Processing (hours)	0:06	0:05	0:07	0:05	0:06	0:15	0:08	0:14	0:06	0:15	0:05	0:15	10:28	0:16	0:06	0:16	0:06	0:16
Total (hours)	**20:25**	19:56	19:20	**19:14**	**20:42**	**56:03**	**19:29**	**48:46**	**19:11**	**42:04**	**19:10**	**57:21**	**31:02**	**34:38**	**22:04**	**38:37**	**20:53**	**48:23**

3. Reanalysis study: 4.2% diagnostic yield

Rady Children's Institute Genomic Medicine

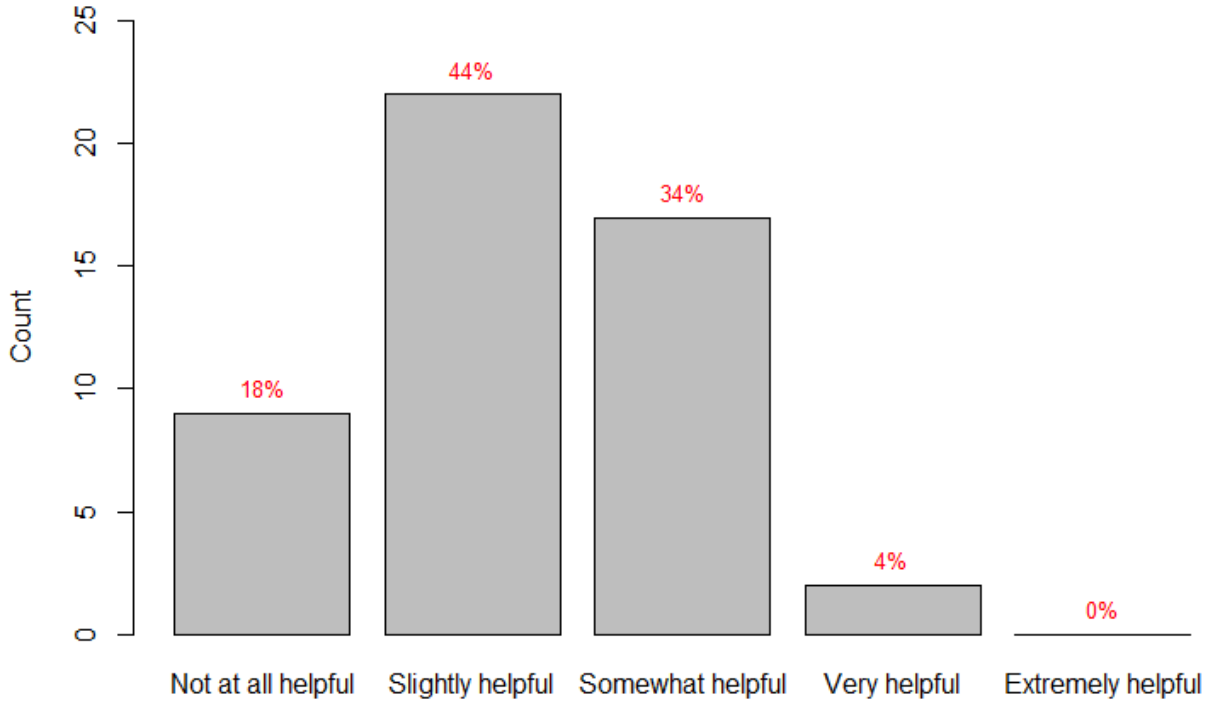
- **Automation of these reanalysis steps reduced the number of variants under consideration by an average of 99.9%.**
- **In two cases, diagnoses were made upon reanalysis, representing a yield of 4.2% (2 of 48).**
- **Four additional cases were flagged with a possible diagnosis to be considered during periodic reanalysis.**
- **An untrained analyst identified these six diagnoses with specificity = 0.83 and sensitivity = 0.76.**

4. Prospective performance: 100% recall

Rady Children's Institute Genomic Medicine

- **Out of 50 patients, the standard diagnostic workflow resulted in 16 (32%) diagnoses**
- **Automated analysis correctly diagnosed all 16 patients (100% recall)**
- **In addition to the standard workflow, analysts found automation to be very helpful in 4% of cases**

"How helpful was automated analysis in addition to the standard workflow?"



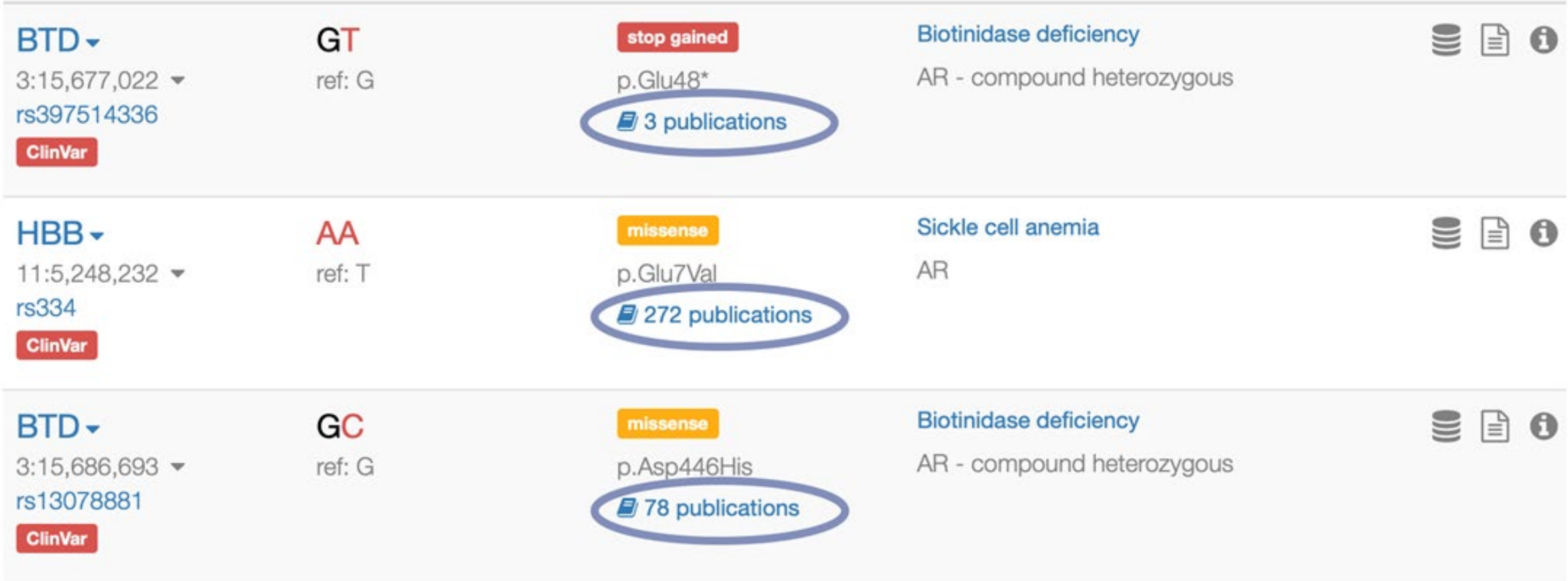
Rady Children's Institute for Genomic Medicine – the clinical lab perspective

Rady Children's Institute Genomic Medicine

- **Hesitation when machine intelligence tools undergo rapid updates**
- **Goes against how clinical lab directors were trained to validate tools**
- **Need sufficient warning prior to updates**
- **Request increased transparency**

Moon's response to requests for transparency

Rady Children's Institute Genomic Medicine

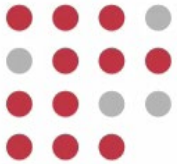


Rady Children's Institute Genomic Medicine

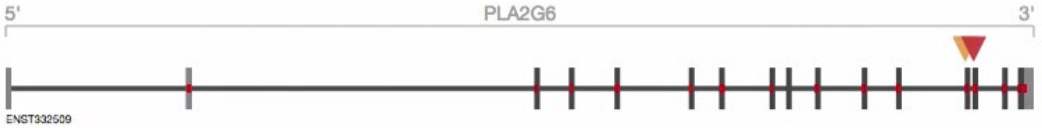
Effect

Transcript	Effect	p.notation	c.notation	Exon rank
ENST332509	missense	p.Arg591Trp	c.1771C>T	13/17
ENST335539	missense	p.Arg537Trp	c.1609C>T	12/16
ENST402064	missense	p.Arg537Trp	c.1609C>T	12/16

Protein prediction



Gene region



Frequency

0.0032% GNOMAD

0 HOMOZYGOTES

1 HETEROZYGOTES

0.0407% DIPLOID

Quality

77 DEPTH

37,40 ALLELE DEPTH

99 GENOTYPE QUALITY

Reported variants

Position	Genotype	Gene	Disorder
22:38511635	G/A	PLA2G6	Infantile neuroaxonal dystrophy 1
22:38512190	G/A	PLA2G6	Infantile neuroaxonal dystrophy 1

Variant discussion

PLA2G6 ▾ 22:38,511,635 ▾	GA ref: G	stop gained p.Arg645*	Infantile neuroaxonal dystrophy 1 AR	

Note

Two variants, a novel stop gained variant and a novel stop gained variant, were detected in heterozygous state in the *PLA2G6* gene (ENST332509: c.1933C>T; p.Arg645* and 332509: c.1933C>T; p.Arg645*). Parental DNA analysis is required to establish a compound heterozygous state of these two variants.

Mutations in *PLA2G6* have been shown to cause Infantile neuroaxonal dystrophy 1 (MIM: 256600), an autosomal recessive condition. The reported clinical phenotype of this patient overlaps with the manifestations of this condition regarding neurodegeneration, developmental regression, nystagmus, spastic tetraplegia, and cerebellar atrophy. The typical age of onset of Infantile neuroaxonal dystrophy 1 ranges from 0 to 10, which is in line with the reported age of onset in this patient (1 y.). Further clinical evaluation of the patient will give more insight into the phenotypic overlap with Infantile neuroaxonal dystrophy 1.

The detected variant causes stop gained. It is absent from gnomAD and absent from dbSNP, but has not previously been associated with disease. Parental DNA analysis (trio analysis) and DNA analysis of other (un)affected relatives, could establish co-segregation of this variant with the reported clinical phenotype.

Classification

Unassigned ▾

References

Enter PubMed ID or PubMed URL

Add publication

The clinical lab perspective continued

Rady Children's Institute Genomic Medicine

- **High sensitivity with automation, but unsure about sensitivity**
  - **Trust will come from large studies from other groups of hundreds of thousands of cases**
- **Development of publically available benchmarks to validate methods after every update**
  - **Example: Genome in a Bottle for clinical validation of genome sequencing**

Conclusion

Rady Children's Institute Genomic Medicine

- **Although the automated diagnostic system is "hands-free", it's supervised at every step by expert bioinformaticians, clinical medical geneticists and clinical lab directors.**
- **May enable effective first-tier, provisional diagnoses or automated re-analysis of unsolved cases**
- **Wide-spread adoption would allow valuable cognitive resources of molecular laboratory directors and analysts to be reserved for difficult cases, manual curation of variants, and clinical report generation**

Acknowledgments

Michelle M. Clark[1], Kiely N. James[1], Amber Hildreth[1,2,3], Sergey Batalov[1], Yan Ding[1], Shimul Chowdhury[1], Kelly Watkins[1], Katarzyna Ellsworth[1], Brandon Camp[1], Cyrielle I. Kint[4], Calum Yacoubian[5], Lauge Farnaes[1,2], Matthew N. Bainbridge[1,6], Curtis Beebe[7], Joshua J. A. Braun[1], Margaret Bray[8], Jeanne Carroll[1,2], Julie A. Cakici[1], Sara A. Caylor[1], Christina Clarke[1], Mitchell P. Creed[9], Jennifer Friedman[1,10], Alison Frith[5], Richard Gain[5], Mary Gaughran[1], Shauna George[7], Sheldon Gilmer[7], Joseph Gleeson[1,10], Jeremy Gore[11], Haiying Grunenwald[12], Raymond L. Hovey[1], Marie L. Janes[1], Kejia Lin[7], Paul D. McDonagh[8], Kyle McBride[7], Patrick Mulrooney[1], Shareef Nahas[1], Daeheon Oh[1], Albert Oriol[7], Laura Puckett[1], Zia Rady[1], Martin G. Reese[13], Julie Ryu[1,2], Lisa Salz[1], Erica Sanford[1,2], Lawrence Stewart[7], Nathaly Sweeney[1,2], Mari Tokita[1], Luca Van Der Kraan[1], Sarah White[1], Kristen Wigby[1,2], Brett Williams[5], Terence Wong[1], Meredith S. Wright[1], Catherine Yamada[1], Peter Schols[4], John Reynders[8], Kevin Hall[12], David Dimmock[1], Narayanan Veeraraghavan[1], Thomas Defay[8], Stephen F. Kingsmore[1*].

[1]Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA;
[2]Department of Pediatrics, University of California San Diego, San Diego, CA 92093, USA;
[3]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA;
[4]Diploid, 3001 Leuven, Belgium;
[5]Clinithink Ltd., London N1 6DR, UK;
[6]Codified Genomics, LLC, Houston, TX 77033, USA;
[7]Rady Children's Hospital, San Diego, CA 92123, USA;
[8]Alexion Pharmaceuticals, Inc., New Haven, CT 06510, USA;
[9]University of Kansas School of Medicine, Kansas City, MO 66160, USA;
[10]Department of Neurosciences, University of California San Diego, San Diego, CA 92093, USA;
[11]Tessella, Needham, MA 02494, USA;
[12]Illumina, Inc., San Diego, CA 92122, USA;
[13]Fabric Genomics, Inc., Oakland, CA 94612, USA.

Support:
NICHD
NHGRI
Illumina
Alexion
Diploid
Clinithink
Fabric